Finding the Projection of a Point onto the Intersection of Convex Sets via Projections onto Halfspaces

Lev M. Bregman[1], Yair Censor[2], Simeon Reich[3]
and Yael Zepkowitz-Malachi[2]

[1]The Institute for Industrial Mathematics,
4 Yehuda Hanakhtom Street, Beer-Sheva 84311, Israel
(bregman@math.bgu.ac.il)

[2]Department of Mathematics,
University of Haifa, Mt. Carmel, Haifa 31905, Israel
({yair,yael}@math.haifa.ac.il)

[3]Department of Mathematics,
The Technion–Israel Institute of Technology,
Haifa 32000, Israel (sreich@techunix.technion.ac.il)

May 9, 2002. Revised: June 8, 2003.

Abstract

We present a modification of Dykstra's algorithm which allows us to avoid projections onto general convex sets. Instead, we calculate projections onto either a halfspace or onto the intersection of two halfspaces. Convergence of the algorithm is established and special choices of the halfspaces are proposed.

The option to project onto halfspaces instead of general convex sets makes the algorithm more practical. The fact that the halfspaces

are quite general enables us to apply the algorithm in a variety of cases and to generalize a number of known projection algorithms.

The problem of projecting a point onto the intersection of closed convex sets receives considerable attention in many areas of mathematics and physics as well as in other fields of science and engineering such as image reconstruction from projections.

In this work we propose a new class of algorithms which allow projection onto certain super halfspaces, i.e., halfspaces which contain the convex sets. Each one of the algorithms that we present gives the user freedom to choose the specific super halfspace from a family of such halfspaces. Since projecting a point onto a halfspace is an easy task to perform, the new algorithms may be more useful in practical situations in which the construction of the super halfspaces themselves is not too difficult.

1 Introduction

The Dykstra algorithm is an iterative procedure which (asymptotically) finds the nearest point projection (also called the orthogonal projection) of any given point onto the intersection of a given family of closed convex sets. It iterates by passing sequentially over the individual sets and projecting onto each one *a deflected version* of the previous iterate. A precise description of the algorithm can be found in Han [35]. The algorithm was first proposed and analyzed by Dykstra [30] in 1983 for a family of closed convex cones in R^n. Boyle and Dykstra [11] studied the algorithm for general convex sets in Hilbert space. Han rediscovered the algorithm in 1988 [35] and investigated its behavior in R^n by using the duality theory of mathematical programming. Han and Lou [36] proposed a simultaneous version of the algorithm in R^n.

Gaffke and Mathar [34] studied the Dykstra algorithm in Hilbert space from a duality standpoint and showed its relation to the method of componentwise cyclic minimization over a Cartesian product. They also proposed a fully simultaneous Dykstra algorithm. Iusem and De Pierro showed in [41] convergence of the simultaneous Dykstra algorithm in both consistent and inconsistent cases in R^n, using Pierra's [51] formalism. Crombez [23] did a similar analysis in Hilbert space. Combettes [18] included the Dykstra algorithm in his review. Bauschke and Borwein [4] analyzed the algorithm for two sets in Hilbert space and generalized the work of Iusem and De Pierro [41] to this setting. Deutsch and Hundal published a rate of convergence study for the polyhedral case in [29], and in [39] established generalizations to an infinite family of sets and to random, rather then cyclic, order control, see also Deutsch [28].

Han [35], as well as Iusem and De Pierro [41], showed that for linear inequality constraints and for linear interval inequality constraints (the polyhedral case), the method of Dykstra becomes the Hildreth algorithm, first published in [38] and studied further by D'Esopo [25], and by Lent and Censor [48].

Censor and Reich [16] proposed a synthesis of Dykstra's algorithm with Bregman distances and obtained a new algorithm that solves the best approximation problem with Bregman projections. However, they established convergence of the resulting Dykstra algorithm with Bregman projections only when the constraints are half-spaces. Bauschke and Lewis [9] provided the first proof for general closed convex constraint sets. Their analysis relies on some strong properties of Bregman distances corresponding to Legendre functions which were treated earlier by Bauschke and Borwein [6]. Bauschke and Lewis [9] also discovered the close relationship between the Dykstra algorithm with Bregman projections and the very general and powerful algorithmic framework of Tseng [54]. Bregman, Censor and Reich show that the Dykstra algorithm with Bregman projections is precisely the nonlinear extension of the Bregman optimization algorithm in [14]. This recognition goes beyond the fact that the two algorithms coincide in the linear constraint case, as was shown by Censor and Reich [16]. It enables these authors to present a new proof and convergence analysis of Dykstra's algorithm with Bregman projections for almost cyclic control sequences, which rests on Bregman's original work [13]. Their work also offers an intuitive geometric interpretation of the iterative steps.

Other algorithms for finding the projection of a point onto the intersection of convex sets are available. Haugazeau proposed such an algorithm in [37] and his ideas were further abstracted and developed by Bauschke and Combettes in [7]. The latter was extended to Bregman projections by Bauschke and Combettes in [8]. Earlier, Pierra discussed yet another method, which is, however, related to Haugazeaue's work, in [50] and [51]. Finally, Combettes constructed a block-iterative outer approximations method in [22].

In the present paper we propose an algorithmic scheme which is a modification of Dykstra's algorithm. It allows us to replace projections onto convex sets by projections onto either a half-space or the intersection of two half-spaces. A method, which replaces projections onto convex sets by projections onto either a half-space or the intersection of two half-spaces, was proposed and studied by Iusem and Svaiter in [42] and in [43], but in a different way. In our work, when we have to project onto the intersection of two half-spaces, our scheme enables us to choose one of the half-spaces from a family of possible half-spaces and the convergence theorem is true for the whole family of possible half-spaces. This feature allows us to construct many specific algorithms within our general scheme and to obtain the result of Iusem and Svaiter (both the algorithm and the convergence theorem) as a special case by making a specific choice of the half-spaces.

2 δ-Super Half-spaces and δ-Super Hyperplanes: Definitions and Construction

The orthogonal projection x' of a point x onto a nonempty closed convex set $E \subseteq R^n$ can be viewed the orthogonal projection of x onto the particular hyperplane H which separates x from E and supports E at x', the closest point to x in E. (For the definitions of a separating hyperplane and a supporting hyperplane consult any book on convex analysis or optimization theory or look, e.g., in Censor and Zenios [17]).

But, of course, at the time of performing such an orthogonal projection neither the point x', nor the separating and supporting hyperplane H are available. In view of the simplicity of performing an orthogonal projection onto a hyperplane, it is natural to ask whether in the construction of iterative projection algorithms one could use other separating supporting hyperplanes, instead of that particular hyperplane H through the closest point to x.

Aside from theoretical interest, this approach leads to algorithms that can be used in practice, provided that the computational effort of finding such other hyperplanes competes favorably with the work involved in performing orthogonal projections directly onto the given sets.

Such an approach was taken by Aharoni, Berman and Censor in [1], where the (δ, η)-algorithm for the convex feasibility problem replaces orthogonal projections onto the convex sets by projections onto separating hyperplanes (see also [17, Algorithm 5.5.1]).

In the present paper we use the rationale behind the (δ, η)-algorithm to deal with another class of mathematical problems, namely, finding the optimal point that minimizes a given objective function over the intersection of given convex sets. We construct a family of half-spaces and hyperplanes with particular properties and replace projections onto convex sets by projections onto a half-space or a hyperplane from this family or onto the intersection of two half-spaces.

In the next subsection we define *δ-super half-spaces* and *δ-super hyperplanes* which play an important role throughout this work.

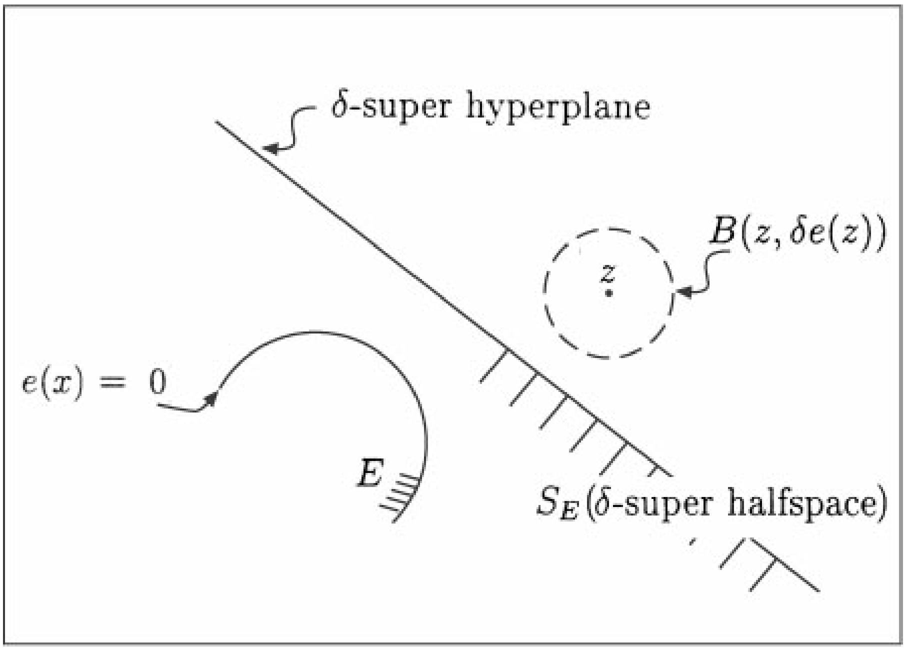

Figure 1: Geometric description of a *δ-Super halfspace* and a *δ-Super hyperplane*

2.1 Definition and Construction of δ-Super Half-spaces and δ-Super Hyperplanes

Let $E \subseteq R^n$ be a nonempty closed convex set defined by $E := \{x \in R^n \mid e(x) \leq 0\}$, where $e\colon\ R^n \to R$ is a convex function, and let $z \in R^n$ be a given point. For $z \notin E$ we wish to construct a half-space which contains the set E, but does not contain any point of the interior of a ball centered at z with radius $\delta e(z)$, for some fixed δ, $0 < \delta \leq 1$. Such a half-space will be called a δ-*Super Half-space* (δ-SHS for short) with respect to the ball described above and E, and its boundary will be called a δ-*Super Hyperplane* (δ-SHP for short). If $z \in E$, then the only possible δ-SHS is defined to be R^n (see Figure 1).

Definition 1 ***(δ-Super Half-space)***
Given are a point $z \in R^n$, a real number δ, $0 < \delta \leq 1$, and a nonempty closed convex set $E := \{x \in R^n \mid e(x) \leq 0\}$, where $e\colon R^n \to R$ is a convex function. For $z \notin E$ define the ball

$$B(z, \delta e(z)) := \{x \in R^n \mid \|x - z\| \leq \delta e(z)\} \tag{1}$$

and a half-space

$$S_E(z) := \{x \in R^n \mid \langle t^E(z), x\rangle \leq \theta_E(z)\}, \tag{2}$$

where $t^E(z) \neq 0$ and $\theta_E(z) \in R$. The set $S_E(z)$ will be called ***a δ-super half-space with respect to*** *$B(z, \delta e(z))$* ***and*** *E if and only if the following two conditions hold:*

$$y \in B(z, \delta e(z)) \text{ implies } \langle t^E(z), y\rangle \geq \theta_E(z), \tag{3}$$

i.e.,

$$S_E(z) \cap \operatorname{int} B(z, \delta e(z)) = \emptyset; \tag{4}$$

and

$$E \subseteq S_E(z). \tag{5}$$

Definition 2 ***(δ-Super Hyperplane)***
If $S_E(z)$ is a δ-SHS, as in Definition 1, and $z \notin E$, then its bounding hyperplane is called ***a δ-super hyperplane (δ-SHP for short) with respect to*** *$B(z, \delta e(z))$* ***and*** *E.* ***If*** *$z \in E$ then the δ-SHS is empty.*

Here is an example of the construction of a δ-SHS. This particular example plays an important role in a special case of our new algorithm (see Section 5.1 below).

Example 3 *Let E be a convex set, $E := \{x \in R^n \mid e(x) \leq 0\}$, where $e(x)$ is a convex function. Let $z \notin E$ and denote by $e'(z)$ any subgradient of e at z. Assume that there exists an $M > 0$ such that $\|e'(z)\| \leq M$ for all z in some bounded set $G \subseteq R^n$. Then*

$$H := \{x \in R^n \mid e(z) + \langle e'(z), x - z\rangle \leq 0\} \tag{6}$$

is a δ-SHS with respect to $B(z, \delta e(z))$ and E for all $z \in G$ and $\delta \leq 1/M$. In order to prove this claim we need to show that (4) and (5) hold. Indeed, let $y \in H$. Then

$$\langle e'(z), y - z\rangle \leq -e(z), \tag{7}$$

by (6). It follows that

$$e(z) \leq |\langle e'(z), y - z\rangle| \leq \|e'(z)\| \cdot \|z - y\|. \tag{8}$$

Using the assumption on the boundedness of the subgradients we obtain

$$e(z) \leq M \cdot \|z - y\|, \tag{9}$$

or

$$e(z) \cdot \frac{1}{M} \leq \|z - y\|. \tag{10}$$

The last inequality shows that whenever δ is chosen such that $\delta \leq 1/M$, we obtain $y \notin$ int $B(z, \delta e(z))$ which implies that $H \cap$ int $B(z, \delta e(z)) = \emptyset$. Next, we show that $E \subseteq H$. Let $x \in E$. Using the well-known subgradient inequality (see, e.g., [52, p. 214]) we have

$$e(x) - e(z) \geq \langle e'(z), x - z\rangle. \tag{11}$$

Since $x \in E$, $e(x) \leq 0$. Thus

$$-e(z) \geq \langle e'(z), x - z\rangle, \tag{12}$$

which implies that $x \in H$, by (6), and as a result of (7)–(12) we obtain that H is indeed a δ-SHS, as claimed.

The last example is very important because there are many algorithms that use projections onto half-spaces of the form of (6), see, e.g., Iusem and Svaiter [42] and Fukushima, [32] and [33]. We will need to make use of the following condition.

Condition 4 *For any E, as in Definition 1, and any bounded set $G \subseteq R^n$, there exists a $\delta \in (0,1)$ such that $z \in G$ and $z \notin E$ imply that the intersection $B(z, \delta e(z)) \cap E$ is empty.*

Remark 5 *Condition 4 is necessary and sufficient to enable us to construct a δ-SHS. In our general algorithmic scheme, which will be presented next, we have to assume that this condition holds. For the special cases of the algorithm, that we treat separately, we show how to choose a δ such that Condition 4 actually holds. Example 3 illustrates such a special case. However, formulating a general sufficient condition for Condition 4 to hold for our general algorithmic scheme still eludes us.*

3 The δ-SHS Algorithm

3.1 The Algorithm

We consider the optimization problem

$$\min\{f(x) \mid x \in Q\}, \tag{13}$$

where $Q = \cap_{i=1}^{m} Q_i$, $Q_i := \{x \in R^n \mid q_i(x) \leq 0\}$, and f, $\{q_i\}_{i=1}^{m}$ are real-valued functions the effective domains of which, dom f and dom q_i, are subsets of R^n. We make the following assumptions regarding the constraints:

- **Assumption (A1).** $q_i(x)$ is convex, $1 \leq i \leq m$.
- **Assumption (A2).** $Q \neq \emptyset$.
- **Assumption (A3).** $Q \cap (\text{dom } f) \neq \emptyset$ and $Q_i \cap \text{int}(\text{dom } f) \neq \emptyset$, for all i , $1 \leq i \leq m$. (But $Q \cap$ int(dom f) may be empty.)
- **Assumption (A4).** The sets dom q_i, $1 \leq i \leq m$, are "wide enough" in the sense that all points appearing in the new Algorithm 8, defined below, belong to int(dom q_i), $1 \leq i \leq m$.

We assume that f is a *Bregman function with zone* $S = \text{int}(\text{dom } f)$ (see the definition of $D_f(x, y)$ in, e.g., [17, Definition 2.1.1]).

Remark 6 *If f is a Bregman function with zone $S = \text{int}(\text{dom } f)$ and if a point $a \in S = \text{int}(\text{dom } f)$ satisfies $\nabla f(a) = 0$, then the problem*

$$\min\{f(x) \mid x \in Q \cap (\text{dom} f)\} \tag{14}$$

achieves its minimum at the Bregman projection with respect to f of the point a onto Q. This follows immediately from the definition of the Bregman projection of a point onto Q, and the fact that $\nabla f(a) = 0$, i.e.,

$$\begin{aligned} P_Q^f(a) &= \arg\min\{D_f(x,a) \mid x \in Q \cap (\text{dom} f)\} \\ &= \arg\min\{f(x) - f(a) - \langle \nabla f(a), x - a\rangle \mid x \in Q \cap \bar{S}\} \\ &= \arg\min\{f(x) \mid x \in Q \cap (\text{dom} f)\}. \end{aligned} \tag{15}$$

We add the next three assumptions on f to make sure that the algorithm is well-defined:

- **Assumption (B1).** The function f is co-finite, which, since it is a Bregman function, implies that the mapping $y = \nabla f(x)$ is a one-to-one mapping of int(dom f) onto R^n (see, e.g., Rockafellar [52, Theorem 26.5]).
- **Assumption (B2).** The function f is *zone consistent* with respect to any half-space and with respect to the intersection of two half-spaces containing points from int(dom f).

Remark 7 *Bauschke and Borwein showed in [6, Theorem 3.14] that if f is a Legendre function, then it is zone consistent. Rockafellar [52, Lemma 26.7] gives a characterization of co-finiteness for differentiable convex functions.*

- **Assumption (B3).** The function f has a global minimum.

The precise description of the new algorithm that we propose for problem (13) is as follows:

Algorithm 8

- ***1. Data at the beginning of the k-th iterative step***

1.1. Current approximation $x^k \in$ int(dom f).

1.2. m vectors $a_i^k \in R^n$ and m real numbers α_i^k, $1 \leq i \leq m$, such that each pair (a_i^k, α_i^k) defines a half-space L_i^k,

$$L_i^k := \{x \in R^n \mid \langle a_i^k, x\rangle \leq \alpha_i^k\}, \tag{16}$$

containing Q_i.

2. Initialization

2.1. x^0 is a (global) minimum point of $f(x)$ on R^n, i.e.,

$$\nabla f(x^0) = 0. \tag{17}$$

2.2. Set $a_i^0 = 0$ and $\alpha_i^0 = 0$ for all $1 \leq i \leq m$.

2.3. Choose a real $\delta \in (0, 1]$ such that Condition 4 holds for the bounded set

$$G = \{x \in R^n \mid D_f(y, x) \leq f(y) - f(x^0)\}, \text{ for some } y \in Q \cap \operatorname{dom} f, \tag{18}$$

and for each Q_i, $i = 1, 2, \ldots, m$. (Note that G is bounded because the partial level sets of the Bregman distance D_f are always bounded, see, e.g., [17, Definition 2.1.1]).

3. Iterative step

3.1. Choose an operating control index $i(k)$ from the ***`almost cyclic control sequence`***. *Recall (see, e.g., [17, Definition 5.1.1]) that an almost cyclic control sequence on $\{1, 2, \ldots, m\}$ is a sequence $\{i(k)\}_{k=0}^{\infty}$ such that $1 \leq i(k) \leq m$ for all $k \geq 0$ and there exists a constant (called the almost cyclicallity constant) $T \geq m$ such that, for all $k \geq 0$,*

$$\{1, 2, \ldots, m\} \subseteq \{i(k + 1), i(k + 2), \ldots, i(k + T)\}. \tag{19}$$

3.2. Calculate $z^k \in R^n$ such that

$$\nabla f(z^k) = \nabla f(x^k) + a_{i(k)}^k. \tag{20}$$

*Such a vector $z^k \in \operatorname{int}(\operatorname{dom} f)$ exists because of Assumption (**B1**).*

3.3. Set x^{k+1}, $\alpha_{i(k)}^{k+1}$, $a_{i(k)}^{k+1}$ by one of the following two possible options:

3.3.1. If $q_{i(k)}(x^k) \leq 0$, then let x^{k+1} be the Bregman projection with respect to f of z^k onto the half-space

$$L_{i(k)}^k := \{x \in R^n \mid \langle a_{i(k)}^k, x\rangle \leq \alpha_{i(k)}^k\}. \tag{21}$$

That is,
if $z^k \in L_{i(k)}^k$, then let $x^{k+1} = z^k$ and define $\lambda_k := 0$,
if $z^k \notin L_{i(k)}^k$, then x^{k+1} and λ_k are calculated from the Karush-Kuhn-Tucker (see, e.g., [49]) conditions

$$\begin{cases} \nabla f(x^{k+1}) = \nabla f(z^k) - \lambda_k a_{i(k)}^k \\ \lambda_k \geq 0, \\ \langle a_{i(k)}^k, x^{k+1}\rangle \leq \alpha_{i(k)}^k, \\ \lambda_k(\langle a_{i(k)}^k, x^{k+1}\rangle - \alpha_{i(k)}^k) = 0. \end{cases} \tag{22}$$

Next set

$$\begin{cases} a_{i(k)}^{k+1} = \lambda_k a_{i(k)}^k, \\ \alpha_{i(k)}^{k+1} = \lambda_k \alpha_{i(k)}^k. \end{cases} \tag{23}$$

3.3.2. If $q_{i(k)}(x^k) > 0$, then let x^{k+1} be the Bregman projection with respect to f of z^k onto the intersection of the following two half-spaces: $L_{i(k)}^k$ which was defined by (21) and

$$S_{Q_{i(k)}}(x^k) := \{x \in R^n \mid \langle t^{i(k)}, x\rangle \leq \theta_{i(k)}\}, \tag{24}$$

which is a δ-SHS with respect to the ball $B(x^k, \delta q_{i(k)}(x^k))$ and $Q_{i(k)}$. In other words,

$$\begin{cases} \langle t^{i(k)}, x\rangle \leq \theta_{i(k)} & \text{for all } x \in Q_{i(k)}, \\ \langle t^{i(k)}, x\rangle \geq \theta_{i(k)} & \text{for all } x \in B(x^k, \delta q_{i(k)}(x^k)), \end{cases} \tag{25}$$

where $t^{i(k)} = t^{i(k)}(x^k) \neq 0$ and $\theta_{i(k)} = \theta_{i(k)}(x^k)$ (see Section 2 for the definition and construction of the δ-SHS). Thus, we calculate the vector x^{k+1}, λ_k and μ_k from the Karush-Kuhn-Tucker *conditions*

$$\begin{cases} \nabla f(x^{k+1}) - \nabla f(z^k) + \lambda_k a^k_{i(k)} + \mu_k t^{i(k)} = 0, \\ \lambda_k \geq 0, \quad \mu_k \geq 0, \\ \langle a^k_{i(k)}, x^{k+1} \rangle \leq \alpha^k_{i(k)}, \\ \langle t^{i(k)}, x^{k+1} \rangle \leq \theta_{i(k)}, \\ \lambda_k (\langle a^k_{i(k)}, x^{k+1} \rangle - \alpha^k_{i(k)}) = 0, \\ \mu_k (\langle t^{i(k)}, x^{k+1} \rangle - \theta_{i(k)}) = 0, \end{cases} \tag{26}$$

and then set

$$\begin{cases} a^{k+1}_{i(k)} = \lambda_k a^k_{i(k)} + \mu_k t^{i(k)}, \\ \alpha^{k+1}_{i(k)} = \lambda_k \alpha^k_{i(k)} + \mu_k \theta_{i(k)}. \end{cases} \tag{27}$$

3.4. For $i \neq i(k)$ do not change a^k_i and α^k_i, i.e., set

$$\begin{aligned} a^{k+1}_i &= a^k_i, \\ \alpha^{k+1}_i &= \alpha^k_i. \end{aligned} \tag{28}$$

Figure 2 describes geometrically the various possibilities of the iterative step of Algorithm 8, in the following way: Cases (i)–(iii) describe iterative steps in which x^k belongs to the convex set $Q_{i(k)}$ (but the modified point z^k can be in $Q_{i(k)}$, in $L^k_{i(k)}$ and not in $Q_{i(k)}$, and not in $L^k_{i(k)}$, respectively). Case (iv) describes the iterative step in which x^k does not belongs to the convex set $Q_{i(k)}$.

Note that the Bregman projection of a point onto a half-space or onto the intersection of two half-spaces exists. According to Assumption (**B2**) the projection belongs to $\operatorname{int}(\operatorname{dom} f)$, thus we have that x^{k+1}, λ_k and μ_k in (22) and (26) exist and $x^{k+1} \in \operatorname{int}(\operatorname{dom} f)$.

Remark 9 *Observe that no relaxation parameters appear in our algorithm. Some of the special cases discussed below employ a sequence $\{\beta_k\}_{k\geq 0}$ of relaxation parameters. Loosely speaking, these parameters overdo or underdo the move prescribed in an iterative step. Relaxation parameters add an extra degree of freedom to the way a method might actually be implemented, and have important consequences for the performance of the method in practice. We,* *however, do have the flexibility to choose a δ-SHS which will lie closer to x^k at each iterate step. In this way, some underrelaxation, i.e., $\epsilon \le \beta_k < 1$ for some arbitrarily small $\epsilon > 0$, can be actually incorporated in our algorithm.*

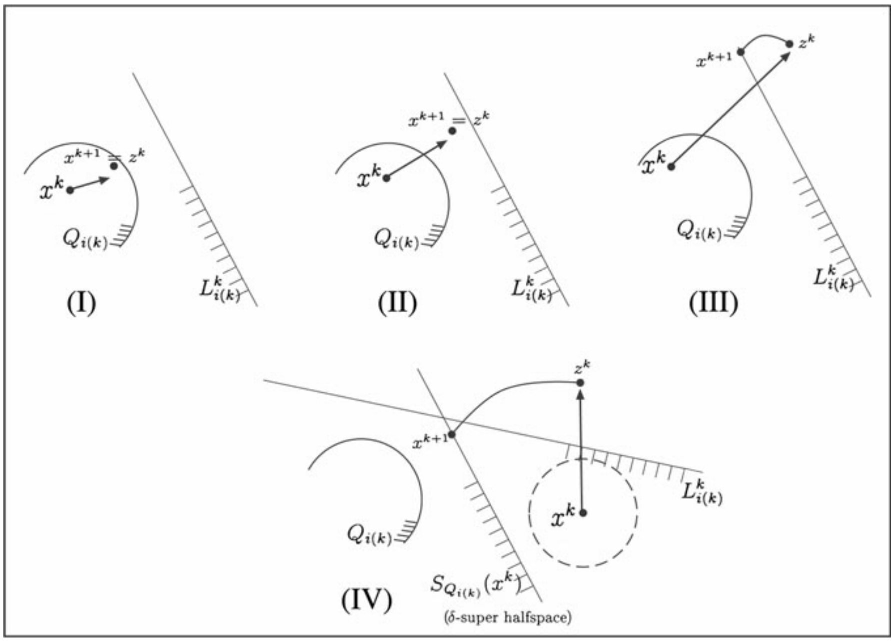

Figure 2: Geometric interpretation of Algorithm 8

The three lemmas below will be used to prove the convergence of Algorithm 8.

Lemma 10 *For any $k \ge 0$, the half-space $L^{k+1}_{i(k)} := \{x \in R^n \mid \langle a^{k+1}_{i(k)}, x\rangle \le \alpha^{k+1}_{i(k)}\}$, defined by the pair $(a^{k+1}_{i(k)}, \alpha^{k+1}_{i(k)})$, generated by Algorithm 8, contains $Q_{i(k)}$.*

Proof. When Step 3.3.1 in Algorithm 8 holds, λ_k is non-negative (see Censor and Zenios [17], Lemma 2.2.2. Notice that there is a sign difference of λ_k between Lemma 2.2.2 in [17] and our Lemma, because of a different definition of the Lagrangian function). Thus, by the definition (23) of $a^{k+1}_{i(k)}$ and $\alpha^{k+1}_{i(k)}$ we obtain $L^{k+1}_{i(k)} = L^k_{i(k)}$, and $L^k_{i(k)}$ contains Q_i (by Step 1.2 of

Algorithm 8). When Step 3.3.2 of Algorithm 8 holds, let $x \in Q_{i(k)}$. We will show that this implies that $x \in L_{i(k)}^{k+1}$. We have

$$\langle a_{i(k)}^{k+1}, x\rangle = \lambda_k \langle a_{i(k)}^k, x\rangle + \mu_k \langle t^{i(k)}, x\rangle \leq \lambda_k \alpha_{i(k)}^k + \mu_k \theta_{i(k)} = \alpha_{i(k)}^{k+1}, \tag{29}$$

where the first equality uses (27), the inequality uses (21) and (24), and the last equality again uses (27). We got that if $x \in Q_{i(k)}$, then $\langle a_{i(k)}^{k+1}, x\rangle \leq \alpha_{i(k)}^{k+1}$, i.e., $x \in L_{i(k)}^{k+1}$. Thus, the lemma is true for Step 3.3.2 too. So Lemma 10 is true in all cases. ■

Lemma 11 *If $\{x^k\}_{k\geq 0}$ and $\{a_i^k\}_{k\geq 0}$ are generated by Algorithm 8, then, for all $k \geq 0$, we have*

$$\nabla f(x^k) + \sum_{i=1}^m a_i^k = 0. \tag{30}$$

Proof. For $k = 0$ the statement is true by (17) and Step 2.2 in the initialization of Algorithm 8. Now we suppose that it is true for some k and we prove it for $k+1$. For the proof we will use the fact (see Step 3.4 of Algorithm 8) that

$$\sum_{\substack{i=1 \\ i\neq i(k)}}^m a_i^{k+1} = \sum_{\substack{i=1 \\ i\neq i(k)}}^m a_i^k. \tag{31}$$

When Step 3.3.1 in Algorithm 8 holds, then using (22), (20), the induction hypothesis, (31) and (23) we obtain

$$\begin{aligned}
\nabla f(x^{k+1}) &= \nabla f(z^k) - \lambda_k a_{i(k)}^k = \nabla f(x^k) + a_{i(k)}^k - \lambda_k a_{i(k)}^k \\
&= (-\sum_{i=1}^m a_i^k) + a_{i(k)}^k - \lambda_k a_{i(k)}^k = (-\sum_{\substack{i=1 \\ i\neq i(k)}}^m a_i^{k+1}) - a_{i(k)}^k + a_{i(k)}^k - \lambda_k a_{i(k)}^k \\
&= (-\sum_{\substack{i=1 \\ i\neq i(k)}}^m a_i^{k+1}) - a_{i(k)}^{k+1} = -\sum_{i=1}^m a_i^{k+1}.
\end{aligned} \tag{32}$$

When Step 3.3.2 in Algorithm 8 holds, then we use (26), (20), the induction hypothesis, (31) and (27), to obtain

$$\begin{aligned}\nabla f(x^{k+1}) &= \nabla f(z^k) - \lambda_k a^k_{i(k)} - \mu_k t^{i(k)} = \nabla f(x^k) + a^k_{i(k)} - \lambda_k a^k_{i(k)} - \mu_k t^{i(k)} \\ &= (-\sum_{i=1}^m a_i^k + a^k_{i(k)}) - \lambda_k a^k_{i(k)} - \mu_k t^{i(k)} \\ &= (-\sum_{\substack{i=1 \\ i\neq i(k)}}^m a_i^{k+1}) - a^k_{i(k)} + a^k_{i(k)} - \lambda_k a^k_{i(k)} - \mu_k t^{i(k)} \\ &= (-\sum_{\substack{i=1 \\ i\neq i(k)}}^m a_i^{k+1}) - a^{k+1}_{i(k)} = -\sum_{i=1}^m a_i^{k+1}. \end{aligned} \tag{33}$$

This completes the proof of Lemma 11. ∎

Lemma 12 *If $\{x^k\}_{k\geq 0}$, $\{a_i^k\}_{k\geq 0}$ and $\{\alpha_i^k\}_{k\geq 0}$ are generated by Algorithm 8, then, for $k \geq 0$, there exists an integer $r = r(k) \geq 0$ and r vectors*

$$y^j \in \{x^0, x^1, \ldots, x^{k-1}\}, \quad 1 \leq j \leq r,$$

such that, for all i, $1 \leq i \leq m$, a_i^k can be represented as a finite linear combination of the normal vectors t^i, generated in Step 3.3.2 of Algorithm 8, at y^j with nonnegative coefficients, that is,

$$a_i^k = \sum_{j=1}^r \gamma_j t^i(y^j), \quad \gamma_j \geq 0, \tag{34}$$

and, for all i, $1 \leq i \leq m$, α_i^k can be represented as

$$\alpha_i^k = \sum_{j=1}^r \gamma_j \theta_i(y^j), \quad \gamma_j \geq 0, \tag{35}$$

where the $\theta_i(y^j)$'s correspond to the $t^i(y^j)$'s generated above.

Proof. For $k = 0$, the left-hand sides of (34) and (35) are zero by Step 2.2 of Algorithm 8. So both statements are true with $\gamma_j = 0$, $j = 1, 2, \ldots, r$. We now assume that the lemma is true for some k and prove it for $k + 1$.

For $i \neq i(k)$ both (34) and (35) hold, because a_i^k and α_i^k do not change, according to Step 3.4 of the algorithm.

For $i = i(k)$, in Step 3.3.1 or in Step 3.3.2 with $\mu_k = 0$, (34) and (35) hold, because, by (23) or (27) and the induction hypothesis,

$$a_{i(k)}^{k+1} = \lambda_k a_{i(k)}^k = \lambda_k(\sum_{j=1}^r \gamma_j t^i(y^j)), \tag{36}$$

and

$$\alpha_{i(k)}^{k+1} = \lambda_k \alpha_{i(k)}^k = \lambda_k(\sum_{j=1}^r \gamma_j \theta_{i(k)}(y^j)). \tag{37}$$

In Step 3.3.2 with $\mu_k > 0$, (34) and (35) hold by (27), the induction hypothesis and the fact that a vector x^k is added to the set $\{x^0, x^1, \dots, x^{k-1}\}$, so we have $t^{i(k)} = t^{i(k)}(y^j)$ for some y^j. Thus we have

$$a_{i(k)}^{k+1} = \lambda_k a_{i(k)}^k + \mu_k t^{i(k)}(x^k) = \lambda_k(\sum_{j=1}^r \gamma_j t^{i(k)}(y^j)) + \mu_k t^{i(k)}(y^j), \tag{38}$$

$$\alpha_{i(k)}^{k+1} = \lambda_k \alpha_{i(k)}^k + \mu_k \theta_{i(k)}(x^k) = \lambda_k(\sum_{j=1}^r \gamma_j \theta_{i(k)}(y^j)) + \mu_k \theta_{i(k)}(y^j). \tag{39}$$

■

4 Convergence of Algorithm 8

We have to make now an additional assumption on the functions q_i. Denote by I_1 the subset of $I = \{1, 2, \dots, m\}$ for which the q_i are affine functions.

- **Assumption (A5).** There exists a point $\bar{y} \in Q \cap \operatorname{dom} f$ such that $q_i(\bar{y}) < 0$ for all $i \in I_2 := I \setminus I_1$.

4.1 The Convergence Theorem

Theorem 13 *Let f be a Bregman function, let q_i, $1 \leq i \leq m$, be convex functions and let Assumptions (**A1**)–(**A5**) and (**B1**)–(**B3**) hold. Then any sequence $\{x^k\}_{k\geq 0}$, generated by Algorithm 8, converges to the solution of (13).*

Proof. The proof is divided into five steps. In Step 1 we define the sequence $\{\varphi_k\}_{k\geq 0}$ by

$$\varphi_k := f(x^k) + \sum_{i=1}^{m}(\langle a_i^k, x^k\rangle - \alpha_i^k) \tag{40}$$

and show that it is increasing. Step 2 proves that

$$\lim_{k\to\infty} D_f(x^{k+1}, x^k) = 0. \tag{41}$$

In step 3 we show that the sequences $\{a_i^k\}_{k\geq 0}$ are bounded for all i, unless $\alpha_i^k - \langle a_i^k, y\rangle = 0$ for all $y \in Q$ and for all $k \geq 0$. We consider the index sets I_1 and I_2 separately and distinguish two possibilities for I_2 . Step 4 shows that the sequence $\{x^k\}_{k\geq 0}$ converges to

$$x^* = \lim_{t\to\infty} x^{k_t} \ , \tag{42}$$

where the sequence $\{x^{k_t}\}_{t\geq 0}$ is defined below, and proves that $x^* \in Q$. In Step 5 we show that for all i, $1 \leq i \leq m$, the limit

$$\lim_{k\to\infty,\ k\in W}(\alpha_i^k - \langle a_i^k, x^*\rangle) = 0 \ , \tag{43}$$

holds for a certain set of indices W (defined in (66)) and from this and other arguments presented there we obtain that Theorem 13 does indeed hold.
Step 1: In order to prove that $\{\varphi_k\}_{k\geq 0}$ is increasing, we show that $\varphi_{k+1} - \varphi_k \geq 0$ for all $k \geq 0$. By definition of φ_k, the fact that for $i \neq i(k)$ we have, by (28), $a_i^{k+1} = a_i^k$ and $\alpha_i^{k+1} = \alpha_i^k$, the definition of D_f and Lemma 11 we

get

$$\begin{aligned}\varphi_{k+1}-\varphi_k &= f(x^{k+1})-f(x^k)+\sum_{i\neq i(k)}\langle a_i^k, x^{k+1}-x^k\rangle\\ &+\langle a_{i(k)}^{k+1}, x^{k+1}\rangle-\alpha_{i(k)}^{k+1}-(\langle a_{i(k)}^k, x^k\rangle-\alpha_{i(k)}^k)\\ &= D_f(x^{k+1},x^k)+\langle\nabla f(x^k), x^{k+1}-x^k\rangle+\sum_{i=1}^m\langle a_i^k, x^{k+1}-x^k\rangle\\ &-\langle a_{i(k)}^k, x^{k+1}-x^k\rangle+\langle a_{i(k)}^{k+1}, x^{k+1}\rangle-\alpha_{i(k)}^{k+1}-(\langle a_{i(k)}^k, x^k\rangle-\alpha_{i(k)}^k)\\ &= D_f(x^{k+1},x^k)+\langle a_{i(k)}^{k+1}, x^{k+1}\rangle-\alpha_{i(k)}^{k+1}\\ &-(\langle a_{i(k)}^k, x^{k+1}\rangle-\alpha_{i(k)}^k). \end{aligned}\tag{44}$$

In order to complete Step 1 we prove the next two assertions. First we show that in all steps of Algorithm 8,

$$\langle a_{i(k)}^{k+1}, x^{k+1}\rangle-\alpha_{i(k)}^{k+1}=0 \tag{45}$$

holds. Indeed in Step 3.3.1, (45) is true by (23) and (21). In Step 3.3.2, we obtain from (27) and (25):

$$\begin{aligned}\langle a_{i(k)}^{k+1}, x^{k+1}\rangle-\alpha_{i(k)}^{k+1} &= \lambda\langle a_{i(k)}^k, x^{k+1}\rangle+\mu\langle t^{i(k)}, x^{k+1}\rangle-\lambda\alpha_{i(k)}^k-\mu\theta_{i(k)}\\ &= \lambda(\langle a_{i(k)}^k, x^{k+1}\rangle-\alpha_{i(k)}^k)+\mu(\langle t^{i(k)}, x^{k+1}\rangle-\theta_{i(k)})\\ &= 0.\end{aligned}\tag{46}$$

Secondly, in all steps of Algorithm 8 we have,

$$\langle a_{i(k)}^k, x^{k+1}\rangle-\alpha_{i(k)}^k\leq 0. \tag{47}$$

This is true because in both Steps 3.3.1 and 3.3.2, $x^{k+1}\in L_{i(k)}^k$. By (44), (45) and (47) we have

$$\varphi_{k+1}-\varphi_k\geq D_f(x^{k+1},x^k)\geq 0, \tag{48}$$

which shows that the sequence $\{\varphi_k\}_{k\geq 0}$ is increasing. We also get, using (44) and (45) in (47) that

$$\varphi_{k+1}-\varphi_k\geq\alpha_{i(k)}^k-\langle a_{i(k)}^k, x^{k+1}\rangle\geq 0. \tag{49}$$

Step 2: Let $y \in \operatorname{dom} f$. Then, by the definition of the Bregman projection with respect to f, Lemma 11 and (40), we have

$$\begin{aligned} D_f(y, x^k) &= f(y) - f(x^k) - \langle \nabla f(x^k), y - x^k \rangle \\ &= f(y) - f(x^k) + \langle \sum_{i=1}^{m} a_i^k, y - x^k \rangle \\ &= f(y) - f(x^k) - \sum_{i=1}^{m} (\langle a_i^k, x^k \rangle - \alpha_i^k) + \sum_{i=1}^{m} (\langle a_i^k, y \rangle - \alpha_i^k) \\ &= f(y) - \varphi_k + \sum_{i=1}^{m} (\langle a_i^k, y \rangle - \alpha_i^k). \end{aligned} \tag{50}$$

Since $y \in Q$, we have that, for all i, $\langle a_i^k, y \rangle - \alpha_i^k \leq 0$, by the definition of L_i^k and (16). Hence, from (50), the last inequality and Step 1,

$$D_f(y, x^k) \leq f(y) - \varphi_k \leq f(y) - \varphi_0 = f(y) - f(x^0), \tag{51}$$

and the sequence $\{x^k\}_{k \geq 0}$ is bounded by the definition of a Bregman function, (see [17, Definition 2.1.1(iv)]). By the left-hand side inequality of (51), we have

$$\varphi_k \leq f(y) \quad \text{for all } y \in Q \cap \operatorname{dom} f. \tag{52}$$

Thus the sequence $\{\varphi_k\}_{k \geq 0}$ is bounded and $\lim_{k \to \infty} \varphi_k$ exists. This fact and (48) imply that (41) holds. Another inequality that follows from (50) is

$$\sum_{i=1}^{m} (\alpha_i^k - \langle a_i^k, y \rangle) \leq f(y) - \varphi_0 \quad \text{for all } y \in Q \cap \operatorname{dom} f. \tag{53}$$

Since $\alpha_i^k - \langle a_i^k, y \rangle \geq 0$ for all i and for all $y \in Q \cap \operatorname{dom} f$, we see that

$$\alpha_i^k - \langle a_i^k, y \rangle \leq f(y) - \varphi_0 \quad \text{for all } i \text{ and for all } y \in Q \cap \operatorname{dom} f. \tag{54}$$

Observe that the derivation of (51) does not depend on δ and could have been reached at even if instead of $S_{Q_{i(k)}}$ in 3.3.2 of Algorithm 8 we would have taken any hyperplane that separates x^k from $Q_{i(k)}$. Therefore, (51) shows that $x^k \in G$, defined by (18), for all $k \geq 0$. Thus, δ is well-defined.

Step 3: This step is divided into three cases.
Step 3(i): Assume that $i \in I_1$ and $\langle b^i, y\rangle - \beta_i = 0$ for all $y \in Q$. In this case we claim that

$$\langle a_i^k, y\rangle - \alpha_i^k = 0 \text{ for all } \ k \geq 0 \text{ and for all } \ y \in Q. \tag{55}$$

To show that (55) holds, recall that for $i \in I_1$, the functions q_i are affine functions, i.e.,

$$q_i(x) = \langle b^i, x\rangle - \beta_i, \quad \partial f(x) = \{b^i\}. \tag{56}$$

By Lemma 12, there exist nonnegative numbers γ_k such that

$$a_i^k = \gamma_k b^i, \quad \alpha_i^k = \gamma_k \beta_i. \tag{57}$$

If $\gamma_k > 0$, then

$$b^i = \frac{a_i^k}{\gamma_k}, \quad \beta_i = \frac{\alpha_i^k}{\gamma_k}. \tag{58}$$

Substituting b^i and β_i of (58) in $\langle b^i, x\rangle - \beta_i = 0$, we get that (55) holds, as claimed. If $\gamma_k = 0$ then $a_i^k = 0$ and $\alpha_i^k = 0$ and (55) holds too.
Step 3(ii): Assume that $i \in I_1$ and $\langle b^i, y\rangle - \beta_i \neq 0$ for some $y \in Q$. In this case the sequences $\{a_i^k\}_{k\geq 0}$ are bounded for all i. To show this, let $\widetilde{y} \in Q \cap \operatorname{dom} f$ satisfy $\beta_i - \langle b^i, \widetilde{y}\rangle = \epsilon > 0$. Using (54) with $y = \widetilde{y}$ we have

$$\alpha_i^k - \langle a_i^k, \widetilde{y}\rangle \leq f(\widetilde{y}) - \varphi_0, \tag{59}$$

and by (57) we obtain

$$\alpha_i^k - \langle a_i^k, \widetilde{y}\rangle = \gamma_k(\beta_i - \langle b^i, \widetilde{y}\rangle) = \gamma_k \epsilon. \tag{60}$$

Hence, by (59) and (60), $\gamma_k \epsilon \leq f(\widetilde{y}) - \varphi_0$, which means that the numbers γ_k are bounded by $(f(\widetilde{y}) - \varphi_0)/\epsilon$, so, by (57), the sequences $\{a_i^k\}_{k\geq 0}$ are indeed bounded.
Step 3(iii): Assume that $i \in I_2$. For all vectors t^i, generated in Step 3.3.2 of Algorithm 8, we have

$$\theta_i \geq \langle t^i, \ y\rangle \ \text{ for all } \ y \in Q_i. \tag{61}$$

Since $\bar{y}$, the existence of which is assumed in Assumption (**A5)**, is an interior point of Q_i, there exists a ball $B(\bar{y},\ \varepsilon_1)$ contained in Q_i. Therefore, $y = \bar{y} + \varepsilon_1 t^i / \|t^i\| \in B(\bar{y},\ \varepsilon_1) \subset Q_i$. Note that $t^i \neq 0$ by (24) and Definition 1. Hence

$$\theta_i - \langle t^i,\ \bar{y}\rangle \geq \langle t^i,\ y\rangle - \langle t^i,\ \bar{y}\rangle = \langle t^i,\ y - \bar{y}\rangle = \varepsilon_1 \left\|t^i\right\|. \tag{62}$$

Using Lemma 12, we have

$$\alpha_i^k - \langle a_i^k, \bar{y}\rangle = \sum_{j=1}^{r} \gamma_j (\theta_i(y^j) - \langle t^i(y^j), \bar{y}\rangle) \geq \varepsilon_i \sum_{j=1}^{r} \gamma_j \left\|t^i(y^i)\right\|. \tag{63}$$

From (54) and the last inequality we obtain

$$f(\bar{y}) - \varphi_0 \geq \varepsilon_i \sum_{j=1}^{r} \gamma_j \left\|t^i(y^i)\right\|, \tag{64}$$

which means that $\sum_{j=1}^{r} \gamma_j \left\|t^i(y^i)\right\|$ is bounded by $(f(\bar{y}) - \varphi_0)/\varepsilon_i$.
From this fact and (34) we have

$$\left\|a_i^k\right\| = \left\|\sum_{j=1}^{r} \gamma_j t^i(y^i)\right\| \leq \sum_{j=1}^{r} \gamma_j \left\|t^i(y^i)\right\| \leq (f(\bar{y}) - \varphi_0)/\varepsilon_i, \tag{65}$$

which proves that the sequences $\{a_i^k\}_{k\geq 0}$ are bounded when $i \in I_2$.
Step 4: From Step 2 we know that the sequence $\{x^k\}_{k\geq 0}$ is bounded, so it must have cluster points. Choose a convergent subsequence $\{x^{k_t}\}_{t\geq 0}$ of $\{x^k\}_{k\geq 0}$ such that $i(k_t) = 1$. Let $\{x^{k_t}\}$ converge to some point x^*. Since $D_f(x^{k+1}, x^k) \to 0$ as $k \to \infty$, the definition of Bregman functions (see [53]) implies that $\{x^{k_t+1}\}_{t\geq 0}$ converges to the point x^*. Repeating this, we get that all the subsequences $\{x^{k_t+1}\}_{t\geq 0}$, $\{x^{k_t+2}\}_{t\geq 0}$, $\ldots$, $\{x^{k_t+T}\}_{t\geq 0}$ converge to x^*, where T is the almost cyclic control constant. Let

$$W = \cup_{t=1}^{\infty} \cup_{j=0}^{T} \{k_t + j\}, \tag{66}$$

i.e., W is the union of the indices belonging to all of the above sequences. It is clear that the sequence $\{x^k\}_{k\in W}$ also converges to x^*. Let us show next that $x^* \in Q$. If Step 3.3.1 appears infinitely many times for $\{x^{k_t}\}_{t\geq 0}$, that is, $q_1(x^{l_t}) \leq 0$ for some subsequence of $\{x^{k_t}\}_{t\geq 0}$, then $q_1(x^*) = \lim_{t\to\infty} q_1(x^{l_t}) \leq 0$.

Hence we have $x^* \in Q_1$ by the definition of Q_1. If Step 3.3.1 appears a finite number of times, then Step 3.3.2 appears infinitely many times. For Step 3.3.2 we know that $x^{k_t+1} \notin \operatorname{int}\big(B(x^{k_t}, \delta q_1(x^{k_t}))\big)$ because x^{k_t+1} is the Bregman projection of z^{k_t} onto the intersection of the two half-spaces $L^{k_t}_{i(k_t)}$ and $S_{Q_{i(k_t)}}$. Therefore,

$$\|x^{k_t+1} - x^{k_t}\| \geq \delta q_1(x^{k_t}), \tag{67}$$

and since $\lim_{t\to\infty} \|x^{k_t+1} - x^{k_t}\| = 0$ (because $\lim_{t\to\infty} x^{k_t+1} = \lim_{t\to\infty} x^{k_t}$) and $\delta > 0$, we have $\lim_{t\to\infty} q_1(x^{k_t}) \leq 0$. Hence $q_1(x^*) = \lim_{t\to\infty} q_1(x^{k_t}) \leq 0$, so $x^* \in Q_1$. Choosing a subsequence $\{x^{k_t+j_t}\}$ $(0 < j_t < T)$ with $i(k_t + j_t) = 2$ which converges to the same point x^*, we see that $x^* \in Q_2$. Repeating this argument for $1 \leq i \leq m$ we obtain $x^* \in Q$.

Step 5: We now show that (43) holds. Take some i, $1 \leq i \leq m$. If $\alpha_i^k - \langle a_i^k, y\rangle = 0$ for all $y \in Q$ and for all k, then (43) is true for this i. Otherwise, we know that $\{a_i^k\}_{k\geq 0}$ is bounded (by Step 3).

If $k \in W$, then W contains a set $W_k = \{p, p+1, \ldots, p+T\}$ containing k. We know that the set $\{i(p), i(p+1), \ldots, i(p+T-1)\}$ contains i (by the choice of the almost cyclic control index). Let $r \in W_k$, $r \leq p + T - 1$, be the nearest integer to k such that $i(r) = i$. We distinguish between two cases according to the values of r and k.

Step 5(i): Assume that $r < k$. In this case $(r+1) \in W_k$ (from the definition of W_k) and $a_i^k = a_i^{r+1}$, $\alpha_i^k = \alpha_i^{r+1}$ since there is no change in a_i^ℓ and α_i^ℓ for $r + 1 \leq \ell \leq k$ (by (28)). Using the last two equations and (45), we obtain

$$\begin{aligned} \alpha_i^k - \langle a_i^k, x^*\rangle &= \alpha_i^{r+1} - \langle a_i^{r+1}, x^*\rangle \\ &= \alpha_{i(r)}^{r+1} - \langle a_i^{r+1}, x^{r+1}\rangle + \langle a_i^{r+1}, x^{r+1} - x^*\rangle \\ &= \langle a_{i(r)}^{r+1}, x^{r+1} - x^*\rangle. \end{aligned} \tag{68}$$

By definition of L_i^k and the last equation, we get

$$0 \leq \alpha_i^k - \langle a_i^k, x^*\rangle \leq \|a_i^{r+1}\| \cdot \|x^{r+1} - x^*\|. \tag{69}$$

Step 5(ii): Assume that $r \geq k$. In this case $a_i^k = a_i^r$ and $\alpha_i^k = \alpha_i^r$ because a_i^ℓ and α_i^ℓ do not change for $k \leq \ell \leq r$ (by 3.4, in the iterative step of Algorithm 8). Hence by the last two equalities and (49),

$$\begin{aligned} \alpha_i^k - \langle a_i^k, x^*\rangle &= \alpha_i^r - \langle a_i^r, x^*\rangle \\ &= \alpha_{i(r)}^r - \langle a_{i(r)}^r, x^{r+1}\rangle + \langle a_i^r, x^{r+1} - x^*\rangle \\ &\leq \varphi_{r+1} - \varphi_r + \langle a_i^r, x^{r+1} - x^*\rangle. \end{aligned} \tag{70}$$

Therefore

$$0 \leq \alpha_i^k - \langle a_i^k, x^* \rangle \leq \|a_i^k\| \cdot \|x^{r+1} - x^*\| + \varphi_{r+1} - \varphi_r. \tag{71}$$

Since r tends to infinity together with k, $r+1 \in W$, the sequences $\{a_i^k\}_{k\geq 0}$ are bounded and $\{\varphi_k\}_{k\geq 0}$ converges, we see that (69) and (71) imply (43). Since (43) holds for all i, we have

$$\lim_{k\to\infty,\ k\in W} \sum_{i=1}^{m} (\alpha_i^k - \langle a_i^k, x^* \rangle) = 0. \tag{72}$$

Applying (50) with $y = x^*$, we get

$$D_f(x^*, x^k) = f(x^*) - \varphi_k + \sum_{i=1}^{m} (\langle a_i^k, x^* \rangle - \alpha_i^k). \tag{73}$$

By the definition of a Bregman function (see [17, Definition 2.1.1]), one has

$$\lim_{k\to\infty,\ k\in W} D_f(x^*, x^k) = 0. \tag{74}$$

Hence (74), (72) and (73) imply that the subsequence $\{\varphi_k\}_{k\geq 0}$, $k \in W$, tends to $f(x^*)$, and since $\lim_{k\to\infty} \varphi_k$ exists,

$$\lim_{k\to\infty} \varphi_k = f(x^*). \tag{75}$$

Since, by (52), $\lim_{k\to\infty} \varphi_k \leq \min\{f(x) \mid x \in Q\}$, we obtain

$$f(x^*) = \min\{f(x) \mid x \in Q\}. \tag{76}$$

From the fact that f is strictly convex (because it is a Bregman function) and has a unique minimum in Q, it follows that the whole sequence $\{x^k\}_{k\geq 0}$ converges to x^*, and the proof is complete. ∎

5 Particular Cases

It is natural to ask, but quite complicated to answer, in what situations all assumptions made in the previous sections hold. We have no simple answer to this question at this time except for the particular cases discussed next. In these cases the choice of the half-spaces is constructively given. First we show that the δ-SHS $S_{Q_{i(k)}}(x^k)$ can be defined via some subgradients at the current point x^k. The second special case deals with the construction of these δ-SHS's, via interior points in the convex sets (using the assumption that in each set we know an interior point). The idea of generating such a case is based on the (δ, η)-algorithm for convex inequalities with interior points (see, e.g., Censor and Zenios [17, Algorithm 5.5.3]).

5.1 Construction of $S_{Q_{i(k)}}(x^k)$ via Subgradient Vectors

We show here a specific choice of the δ-SHS's $S_{Q_{i(k)}}(x^k)$, made by constructing each of the δ-SHS's via subgradients. In this case we use underrelaxation parameters to define $S_{Q_{i(k)}}(x^k)$.

Let $\{\beta_k\}_{k\geq 0}$ be an infinite sequence of underrelaxation parameters such that $0 < \epsilon \leq \beta_k \leq 1$ for all $k \geq 0$, with some arbitrarily small given ϵ. Let $v^{i(k)} = v^{i(k)}(x^k)$ be a subgradient of $q_{i(k)}$ at x^k.

Theorem 14 *Assume that $\partial q_i(G)$ is bounded for any bounded subset $G \subseteq$ $\operatorname{dom} q_i$, for all $i = 1, 2, \ldots, m$. If in Algorithm 8 one uses, for all $k \geq 0$,*

$$0 \neq t^{i(k)}(x^k) = v^{i(k)} \in \partial q_{i(k)}(x^k), \tag{77}$$

and

$$\theta_{i(k)} = \langle v^{i(k)}, x^k \rangle - \beta_k q_{i(k)}(x^k), \tag{78}$$

to construct $S_{Q_{i(k)}}(x^k)$ by (24) whenever $q_{i(k)}(x^k) > 0$, then $S_{Q_{i(k)}}(x^k)$ is a δ-SHS.

Proof. Let $S_{Q_{i(k)}}(x^k)$ be the half-space defined by (77) and (78), i.e.,

$$S_{Q_{i(k)}}(x^k) = \{x \in R^n \mid \langle v^{i(k)}, x - x^k \rangle \leq -\beta_k q_{i(k)}(x^k)\}. \tag{79}$$

In order to conclude that the half-space $S_{Q_{i(k)}}(x^k)$ is a δ-SHS, we must show that

$$S_{Q_{i(k)}}(x^k) \cap \operatorname{int} B(x^k, \delta q_{i(k)}(x^k)) = \emptyset \tag{80}$$

and, by Definition 1, that

$$S_{Q_{i(k)}}(x^k) \supseteq Q_{i(k)}. \tag{81}$$

First we show that (80) holds. In Step 3.3.2 of Algorithm 8, x^{k+1} is the Bregman projection of z^k onto the intersection $L^k_{i(k)} \cap S_{Q_{i(k)}}(x^k)$. This implies that $x^{k+1} \in S_{Q_{i(k)}}(x^k)$, that is,

$$\langle v^{i(k)}, x^{k+1} - x^k \rangle \leq -\beta_k q_{i(k)}(x^k). \tag{82}$$

It follows that

$$\beta_k q_{i(k)}(x^k) \leq \mid \langle v^{i(k)}, x^k - x^{k+1} \rangle \mid \leq \| v^{i(k)} \| \cdot \| x^k - x^{k+1} \|. \tag{83}$$

According to the comment in the last sentence of Step 2 of the proof, x^k belongs to the set G, defined by (18). Using the assumption on the boundedness of the subgradients, we have $\| v^{i(k)} \| \leq M$. Hence

$$\beta_k q_{i(k)}(x^k) \leq M \| x^k - x^{k+1} \|, \tag{84}$$

which implies that

$$\frac{1}{M} \beta_k q_{i(k)}(x^k) \leq \| x^k - x^{k+1} \|. \tag{85}$$

Taking any

$$\delta \leq \inf \{ \frac{1}{M} \beta_k \mid k \geq 0 \} = \frac{\epsilon}{M}, \tag{86}$$

we get

$$x^{k+1} \notin \operatorname{int} B(x^k, \delta q_{i(k)}(x^k)) \text{ for all } \ k \geq 0, \tag{87}$$

which implies that (80) is true. We now show that (81) also holds. Let $x \in Q_{i(k)}$, i.e., $q_{i(k)}(x) \leq 0$. By the *subgradient inequality* we have

$$q_{i(k)}(x) - q_{i(k)}(x^k) \geq \langle v^{i(k)}, x - x^k \rangle. \tag{88}$$

Thus

$$-q_{i(k)}(x^k) \geq \langle v^{i(k)}, x - x^k \rangle. \tag{89}$$

Since $q_{i(k)}(x^k) > 0$, both sides of (89) are negative. Hence

$$-\beta_k q_{i(k)}(x^k) \geq \langle v^{i(k)}, x - x^k \rangle, \tag{90}$$

i.e., $x \in S_{Q_{i(k)}}(x^k)$ by (79), which implies that (81) does indeed hold. This completes the proof. ■

5.2 Construction of $S_{Q_{i(k)}}(x^k)$ via Interior Points in the Sets

In Case 3.3.2 of Algorithm 8 we can construct the half-spaces $S_{Q_{i(k)}}(x^k)$ for (24) by still another method.

- **Assumption (C).** There are m given interior points $y^i \in \operatorname{int} Q_i$, $1 \leq i \leq m$.

Method for the construction of the half-spaces $S_{Q_{i(k)}}(x^k)$ by (24): If $q_{i(k)}(x^k) > 0$ (i.e., we are in Case 3.3.2 of Algorithm 8), choose some $0 \leq h \leq 1$, define

$$\bar{x}(h) = hy^{i(k)} + (1-h)x^k, \tag{91}$$

and solve the nonlinear equation

$$q_{i(k)}(\bar{x}(h)) = 0. \tag{92}$$

Denote by h_k the smallest value of h for which $\bar{x}(h)$ solves (92) and set

$$\bar{x}^k = \bar{x}(h_k). \tag{93}$$

Then calculate a subgradient

$$t^{i(k)} \in \partial q_{i(k)}(\bar{x}^k). \tag{94}$$

(If $q_{i(k)}$ is differentiable at $(\bar{x}^k)$ then $t^{i(k)} = \nabla q_{i(k)}(\bar{x}^k))$ and

$$\theta_{i(k)} = \langle t^{i(k)}, \bar{x}^k \rangle, \tag{95}$$

and define $S_{Q_{i(k)}}(x^k)$ by

$$S_{Q_{i(k)}}(x^k) = \{x \in R^n \mid \langle t^{i(k)}, x \rangle \leq \langle t^{i(k)}, \bar{x}^k \rangle\}. \tag{96}$$

See Figure 3 for a geometric description of the construction of $S_{Q_{i(k)}}(x^k)$ via interior points in the sets.

Theorem 15 *Under Assumption (C), whenever $q_{i(k)}(x^k) > 0$ in Algorithm 8 and we use the method described above to construct $S_{Q_{i(k)}}(x^k)$, then there exists a $\delta > 0$ such that $S_{Q_{i(k)}}(x^k)$ is a δ-SHS.*

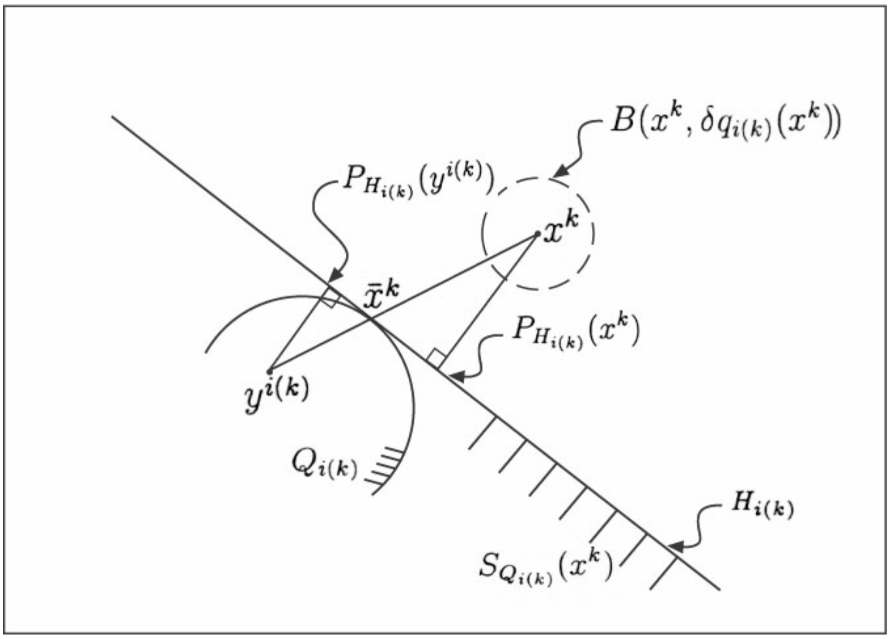

Figure 3: Geometric description of the construction of $S_{Q_{i(k)}}(x^k)$ via interior points in the sets.

Proof. We first show that (81) holds. If $x \in Q_{i(k)}$, then $q_{i(k)}(x) \leq 0$. Using the subgradient inequality

$$\langle t^{i(k)}, x - \bar{x}^k \rangle \leq q_{i(k)}(x) - q_{i(k)}(\bar{x}^k), \tag{97}$$

and the fact that $q_{i(k)}(x) \leq 0$ and $q_{i(k)}(\bar{x}^k) = 0$, we obtain

$$\langle t^{i(k)}, x - \bar{x}^k \rangle \leq 0. \tag{98}$$

In other words,

$$\langle t^{i(k)}, x \rangle \leq \langle t^{i(k)}, \bar{x}^k \rangle. \tag{99}$$

Thus $x \in S_{Q_{i(k)}}(x^k)$, by (96). We show now that (80) holds. Let

$$H_{i(k)} = \{x \in R^n \mid \langle t^{i(k)}, x \rangle = \langle t^{i(k)}, \bar{x}^k \rangle\}. \tag{100}$$

Since $x^k \notin Q_{i(k)}$ (otherwise we do not use (96)), we have, by a simple geometric consideration,

$$\|P_{H_{i(k)}}(x^k) - x^k\| = \frac{\|x^k - \bar{x}^k\| \cdot \|y^{i(k)} - P_{H_{i(k)}}(y^{i(k)})\|}{\|y^{i(k)} - \bar{x}^k\|}, \tag{101}$$

where $P_{H_{i(k)}}(x^k)$ is the orthogonal projection of x^k onto $H_{i(k)}$ (see also [17, Figure 5.8].) Since $\{x^k\}_{k\geq 0}$ is bounded (see the comment in the last sentence of Step 2), (101) implies that there is a positive M such that for all $k \geq 0$,

$$\|y^{i(k)} - \bar{x}^k\| \leq M. \tag{102}$$

By Bauschke and Borwein [5, Proposition 7.8 and Corollary 7.9], $q_{i(k)}$ is Lipschitz continuous. Hence there is a positive L such that

$$|q_{i(k)}(x^k) - q_{i(k)}(\bar{x}^k)| \leq L\|x^k - \bar{x}^k\|. \tag{103}$$

Since $q_{i(k)}(\bar{x}^k) = 0$ by (92) and (93) , we obtain

$$q_{i(k)}(x^k) \leq L\|x^k - \bar{x}^k\|. \tag{104}$$

We also have

$$\|y^{i(k)} - P_{H_{i(k)}}(y^{i(k)})\| \geq d(y^{i(k)}, bd\ Q_{i(k)}) \geq d > 0, \tag{105}$$

where

$$d := \min\{d(y^i, bd\ Q_i) \mid 1 \leq i \leq m\}. \tag{106}$$

It follows from (101), (102), (104) and (105) that

$$\|P_{H_{i(k)}}(x^k) - x^k\| \geq \frac{q_{i(k)}(x^k) \cdot d}{M \cdot L}. \tag{107}$$

Let $x^{k+1} \in S_{Q_{i(k)}}(x^k)$. Then we also obtain

$$\|x^{k+1} - x^k\| \geq \|P_{H_{i(k)}}(x^k) - x^k\| \geq \frac{q_{i(k)}(x^k) \cdot d}{M \cdot L}. \tag{108}$$

Taking $\delta \leq \frac{d}{M \cdot L}$, we have that

$$x^{k+1} \notin \operatorname{int} B(x^k, \delta q_{i(k)}(x^k)) \text{ for all } k \geq 0. \tag{109}$$

This completes the proof of Theorem 15.

■

Acknowledgement. We thank an anonymous referee for his insightful and constructive report which led to a much improved paper. Part of this work was done at the Center for Computational Mathematics and Scientific Computation (CCMSC) at the University of Haifa and was supported by research grant 592/00 from the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities. S. Reich's work was also partially supported by the Fund for the promotion of Research at the Technion and by the Technion VPR Fund.

References

[1] R. Aharoni, A. Berman and Y. Censor, An interior points algorithm for the convex feasibility problem, *Advances in Applied Mathematics* **4** (1983) 479–489.

[2] M. Avriel, *Nonlinear Programing: Analysis and Methods.* Prentice-Hall, Englewood Cliffs, NJ, USA, 1976.

[3] H. H. Bauschke, The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space, *Journal of Mathematical Analysis and Applications* **202** (1996) 150–159.

[4] H. H. Bauschke and J. M. Borwein, Dykstra's alternating projection algorithm for two sets, *Journal of Approximation Theory* **79** (1994) 418–443.

[5] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review* **38** (1996) 367–426.

[6] H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *Journal of Convex Analysis* **4** (1997) 27–67.

[7] H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Mathematics of Operations Research* **26** (2001) 248–264.

[8] H. H. Bauschke and P. L. Combettes, Construction of best Bregman approximations in reflexive Banach spaces, *Proceedings of the American Mathematical Society*, to appear. (available at: http://www.ann.jussieu.fr/~plc/publist.html).

[9] H. H. Bauschke and A. S. Lewis, Dykstra's algorithm with Bregman projections: a convergence proof, *Optimization* **48** (2000) 409–427.

[10] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[11] J. P. Boyle and R. L. Dykstra, A method of finding projections onto the intersection of convex sets in Hilbert spaces. in: *Advances in Order Restricted Statistical Inference*, Lecture Notes in Statistics, Springer-Verlag, Berlin, Germany, Volume **37** (1986) 28–47.

[12] L. M. Bregman, The method of successive projection for finding a common point of convex sets, *Soviet Mathematics Dokl.* **6** (1965) 688–692.

[13] L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics* **7** (1967) 200–217.

[14] L. M. Bregman, Y. Censor and S. Reich, Dykstra's algorithm as the nonlinear extension of Bregman's optimization method, *Journal of Convex Analysis* **6** (1999) 319–333.

[15] Y. Censor and A. Lent, An iterative row-action method for interval convex programming, *Journal of Optimization Theory and Applications* **34** (1981) 321–353.

[16] Y. Censor and S. Reich, The Dykstra algorithm with Bregman projections, *Communications in Applied Analysis* **2** (1998) 407–419.

[17] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications.* Oxford University Press, New York, NY, USA, 1997.

[18] P. L. Combettes, Signal recovery by best feasible approximation, *IEEE Transactions on Image Processing* **IP-2** (1993) 269–271.

[19] P. L. Combettes, The foundations of set theoretic estimation, *Proceedings of the IEEE* **81** (1993) 182–208.

[20] P. L. Combettes, Constrained image recovery in a product space, *Proceedings of the 1995 International Conference on image Processing* (1995) 25–28.

[21] P. L. Combettes, Construction d'une point fixe commun à une famille de contractions fermes, *Comptes Rendus de l'Académie des Sciences de Paris, Série I,* **320** (1995) 1385–1390.

[22] P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM Journal on Control and Optimization* **38** (2000) 538–565.

[23] G. Crombez, Finding projections onto the intersection of convex sets in Hilbert spaces, *Numerical Functional Analysis and Optimization* **16** (1995) 637–652.

[24] I. Csiszár, Why least squares and maximal entropy? An axiomatic approach to inference for linear inverse problems, *The Annals of Statistics* **19** (1991) 2032–2066.

[25] D. A. D'Esopo, A convex programming procedure, *Naval Research Logistics Quarterly* **6** (1959) 33–42.

[26] A. R. De Pierro and A. N. Iusem, A relaxed version of Bregman's method for convex programming, *Journal of Optimization Theory and Applications* **51** (1986) 421–440.

[27] F. Deutsch, The method of alternating orthogonal projections, in: *Approximation Theory, Spline Functions and Application*, S. P. Singh (editor), Kluwer Academic Publishers, Amsterdam, The Netherlands, 1992, 105–121.

[28] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer-Verlag, New York, NY, USA, 2001.

[29] F. Deutsch and H. Hundal, The rate of convergence of Dykstra's cyclic projections algorithm: The polyhedral case, *Numerical Functional Analysis and Optimization* **15** (1994) 537–565.

[30] R. L. Dykstra, An algorithm for restricted least squares regression, *Journal of the American Statistical Association* **78** (1983) 837–842.

[31] R. L. Dykstra, An iterative procedure for obtaining I-projections onto the intersection of convex sets, *The Annals of Probability* **13** (1985) 975–984.

[32] M. Fukushima, An outer approximation algorithm for solving general convex programs, *Operations Research* **31** (1983) 101–113.

[33] M. Fukushima, A relaxed projection method for variational inequalities, *Mathematical Programing* **35** (1986) 58–70.

[34] N. Gaffke and R. Mathar, A cyclic projection algorithm via duality, *Metrika* **36** (1989) 29–54.

[35] S.-P. Han. A successive projection method, *Mathematical Programming* **40** (1988) 1–14.

[36] S.-P. Han and G. Lou, A parallel algorithm for a class of convex programs, *SIAM Journal on Control and Optimization* **26** (1988) 345–355.

[37] Y. Haugazeau, Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes, *Ph. D. Thesis, Université de Paris*, Paris, France, 1968.

[38] C. Hildreth, A quadratic programming procedure, *Naval Research Logistics Quarterly* **4** (1957) 79–85. Erratum, ibid., p. 361.

[39] H. Hundal and F. Deutsch, Two generalizations of Dykstra's cyclic projections algorithm, *Mathematical Programming* **77** (1997) 335–355.

[40] A. N. Iusem and A. R. De Pierro, A simultaneous iterative method for computing projections on polyhedra, *SIAM Journal on Control and Optimization* **25** (1987) 231–243.

[41] A. N. Iusem and A. R. De Pierro, On the convergence of Han's method for convex programming with quadratic objective. *Mathematical Programming* **52** (1991) 265–284.

[42] A. N. Iusem and B. F. Svaiter, A row-action method for convex programming, *Mathematical Programming* **64** (1994) 149–171.

[43] A. N. Iusem and B. F. Svaiter, Primal-dual row-action method for convex programming, *Journal of Optimization Theory and Applications* **86** (1995) 73–112.

[44] K. C. Kiwiel and B. Lopuch, Surrogate projection methods for finding fixed points of firmly nonexpansive mappings, *SIAM Journal on Optimization* **77** (1997) 1084–1102.

[45] K. C. Kiwiel, Block-iterative surrogate projection method for convex feasibility problems, *Linear Algebra and Its Applications* **215** (1995) 225–260.

[46] K. C. Kiwiel, Free-steering relaxation methods for problems with strictly convex sets and linear constraints, *Mathematics of Operations Research* **22** (1997) 326–349.

[47] P. L. Lions, Approximation de points fixes de contractions, *Comptes Rendus de l'Académie des Sciences de Paris* **284**. Série A (1977) 1357–1359.

[48] A. Lent and Y. Censor, Extensions of Hildreth's row-action method for quadratic programming, *SIAM Journal on Control and Optimization* **18** (1980) 444–454.

[49] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley and sons, New York, NY, USA, 1969.

[50] G. Pierra, Eclatement de contraintes en parallèle pour la minimisation d'une forme quadratique, in: *Lecture Notes in Computer Science*, Vol. **41**, Springer-Verlag, New York, NY, USA, 1976, pp. 96–115.

[51] G. Pierra, Decomposition through formalization in a product space, *Mathematical Programming* **28** (1984) 96–115.

[52] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, USA, 1970.

[53] M. V. Solodov and B. F. Svaiter, An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, *Mathematics of Operations Research* **25** (2000) 214–230.

[54] P. Tseng, Dual coordinate ascent methods for non-strictly convex minimization, *Mathematical Programming* **59** (1993) 231–247.